

# 光电工程

## Opto-Electronic Engineering

中文核心期刊 中国科技核心期刊  
Scopus CSCD

### OCT图像多教师知识蒸馏超分辨率重建

陈明惠, 芦焱琦, 杨文逸, 王援柱, 邵怡

#### 引用本文:

陈明惠, 芦焱琦, 杨文逸, 等. OCT图像多教师知识蒸馏超分辨率重建[J]. *光电工程*, 2024, 51(7): 240114.

Chen M H, Lu Y Q, Yang W Y, et al. Super-resolution reconstruction of retinal OCT image using multi-teacher knowledge distillation network[J]. *Opto-Electron Eng*, 2024, 51(7): 240114.

<https://doi.org/10.12086/oe.2024.240114>

收稿日期: 2024-05-15; 修改日期: 2024-08-08; 录用日期: 2024-08-09

### 相关论文

#### 引入双编码器模型的OCT视网膜图像分割

陈明惠, 王腾, 袁媛, 柯舒婷

*光电工程* 2023, 50(10): 230146 doi: 10.12086/oe.2023.230146

#### 线聚焦谱域光学相干层析成像的分段色散补偿像质优化方法

杨建文, 黄江杰, 何益, 史国华

*光电工程* 2024, 51(6): 240042 doi: 10.12086/oe.2024.240042

#### 光学相干层析技术在眼科手术导航方面的研究进展

李云耀, 樊金宇, 蒋天亮, 唐宁, 史国华

*光电工程* 2023, 50(1): 220027 doi: 10.12086/oe.2023.220027

#### Deblurring, artifact-free optical coherence tomography with deconvolution-random phase modulation

Xin Ge, Si Chen, Kan Lin, Guangming Ni, En Bo, Lulu Wang, Linbo Liu

*Opto-Electronic Science* 2024, 3(1): 230020 doi: 10.29026/oes.2024.230020

更多相关论文见光电期刊集群网站 



<http://cn.ojournal.org/oe>



OE\_Journal



Website

DOI: 10.12086/oe.2024.240114

# OCT 图像多教师知识蒸馏超分辨率重建

陈明惠<sup>1\*</sup>, 芦焱琦<sup>1</sup>, 杨文逸<sup>1</sup>, 王援柱<sup>2</sup>, 邵 怡<sup>3</sup>

<sup>1</sup>上海理工大学健康科学与工程学院, 上海介入医疗器械工程技术研究中心, 教育部医学光学工程中心, 上海 200093;

<sup>2</sup>上海瑞柯恩激光技术有限公司, 上海 200120;

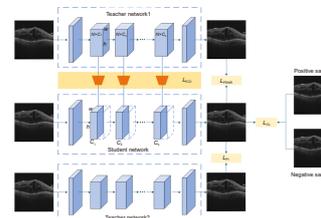
<sup>3</sup>上海市第一人民医院, 上海 200080

**摘要:** 光学相干断层成像 (OCT) 广泛应用于眼科诊断与辅助治疗, 但其成像质量不可避免地受到散斑噪声和运动伪影影响。本文提出了一种针对 OCT 超分辨率任务的多教师知识蒸馏网络 MK-OCT, 使用不同优势的教师网络训练平衡、轻量级和高效的学生网络。MK-OCT 中高效通道蒸馏方法 ECD 的使用也使得模型能够更好地保留视网膜图像的纹理信息, 满足临床需要。实验结果表明, 与经典超分辨率网络相比, 本文所提模型在重建精度和感知质量两个方面均表现优异, 模型尺寸更小, 计算量更少。

**关键词:** 医学图像; 光学相干断层图像; 超分辨率; 知识蒸馏; 对比学习

**中图分类号:** TP391

**文献标志码:** A



陈明惠, 芦焱琦, 杨文逸, 等. OCT 图像多教师知识蒸馏超分辨率重建 [J]. 光电工程, 2024, 51(7): 240114

Chen M H, Lu Y Q, Yang W Y, et al. Super-resolution reconstruction of retinal OCT image using multi-teacher knowledge distillation network[J]. *Opto-Electron Eng*, 2024, 51(7): 240114

## Super-resolution reconstruction of retinal OCT image using multi-teacher knowledge distillation network

Chen Minghui<sup>1\*</sup>, Lu Yanqi<sup>1</sup>, Yang Wenyi<sup>1</sup>, Wang Yuanzhu<sup>2</sup>, Shao Yi<sup>3</sup>

<sup>1</sup>Shanghai Engineering Research Center of Interventional Medical, Shanghai Institute for Interventional Medical Devices, School of Health Sciences and Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China;

<sup>2</sup>Shanghai Raykeen Laser Technology Co., Ltd., Shanghai 200120, China;

<sup>3</sup>Shanghai General Hospital, Shanghai 200080, China

**Abstract:** Optical coherence tomography (OCT) is widely used in ophthalmic diagnosis and adjuvant therapy, but its imaging quality is inevitably affected by speckle noise and motion artifacts. This article proposes a multi teacher knowledge distillation network MK-OCT for OCT super-resolution tasks, which uses teacher networks with different advantages to train balanced, lightweight, and efficient student networks. The use of efficient channel distillation

收稿日期: 2024-05-15; 修回日期: 2024-08-08; 录用日期: 2024-08-09

基金项目: 上海市科委产学研医项目 (15DZ1940400)

\*通信作者: 陈明惠, cmhui.43@163.com。

版权所有©2024 中国科学院光电技术研究所

method ECD in MK-OCT also enables the model to better preserve the texture information of retinal images, meeting clinical needs. The experimental results show that compared with classical super-resolution networks, the model proposed in this paper performs well in both reconstruction accuracy and perceptual quality, with smaller model size and less computational complexity.

**Keywords:** medical images; optical coherence tomography images; super-resolution; knowledge distillation; contrastive learning

## 1 引言

临床上, 以光学相干断层成像 (又名光学相干层析成像, optical coherence tomography, OCT) 技术获取的视网膜图像是眼部疾病诊断和治疗的重要依据。OCT 技术利用光的低相干特性, 获得入射光在眼部不同深度的散射和反射信号进行成像, 眼科医生通过 OCT 图像可以判断分析患者视网膜的病理结构和变化, 有助于早期发现和治疗疾病。随着 OCT 技术在眼科疾病预防、诊断、治疗方面地不断深入, 医生对于 OCT 视网膜图像的质量要求也随之提高。在 OCT 成像过程中, 测量光束在进入组织内部时发生的散射会使得 OCT 图像中出现散斑噪声, 导致图像中重要的细节信息被掩盖。另外, 采集图像过程中被检测者无意识的身体运动等也会导致 OCT 图像中存在伪影, 影响临床诊断并干扰对图像的后续自动化分析效果。因此去除 OCT 视网膜图像中的噪声、提高视网膜图像的分辨率一直以来都是 OCT 图像处理领域的热点问题<sup>[1]</sup>。

随着深度学习的发展, 研究者们提出了许多去除 OCT 图像噪声、提高分辨率的网络。2019 年, Huang 等人<sup>[2]</sup>将深度学习用于 OCT 图像的超分辨率任务, 通过对超分辨率生成对抗网络框架的改进, 提出了可以同时去除 OCT 图像噪声和超分辨率重建的模型 (network for simultaneously denoise and super-resolve OCT images, SDSR-OCT)。与基于稀疏性的方法相比, SDSR-OCT 在客观指标上显示出很好的性能。然而, SDSR-OCT 实验中仅使用了 10 对低分辨率 (low resolution, LR)-高分辨率 (high resolution, HR) 图像作为训练集对模型进行训练, 尽管通过切分图像进行了数据扩充, 然而对于视网膜图像来说, 图像上下部分存在大量没有生理结构的空白部分, 因此该方法存在一定缺陷, 泛化性能有待验证, 可能无法可靠地重建视网膜的结构特征。Das 等人<sup>[3]</sup>提出了一种基于无监督生成对抗网络 (generative adversarial network, GAN) 的 SR 框架, 使用具有周期一致性和身份映射

先验的对抗性学习来进行 OCT 图像去噪和超分辨率。Qiu 等人<sup>[4]</sup>采用两个上下采样网络, 即半监督 U-Net (U-Net (semi)) 和半监督 DBPN (deep back projection networks, DBPN (semi)) 来研究 OCT 图像去噪和超分辨率重建的有效性。本课题组为了解决重建图像边缘细节缺失的问题, 弥补基于卷积神经网络的 OCT 超分辨率重建网络存在的不足, 提出了一种基于金字塔长程 Transformer 的 OCT 图像超分辨率网络, 利用 Transformer 的全局特征提取能力对 OCT 图像进行重建, 在重建精度的客观指标上取得了比较好的成果<sup>[5]</sup>。在此过程中, 本课题组发现了 OCT 超分辨率网络始终存在的一个问题, 即重建图像在重建精度和感知质量上的不平衡。

总结来看, OCT 超分辨率网络可以大致分成两类: 面向重建精度<sup>[2-4]</sup>和面向感知质量<sup>[6-7]</sup>。其中面向重建精度的网络大多数是通过均方误差 (mean squared error, MSE) 进行优化, MSE 用于测量超分辨率图像 (super resolution, SR) 与真实图像 HR 之间的逐像素距离, 使得生成的 SR 图像具有更高的重建精度和结构一致性度量值<sup>[8]</sup>。然而, 测量像素误差很难捕捉到图像之间的感知差异, 由此生成的 SR 图像虽然可以在峰值信噪比上得到较好的结果, 但可能存在模糊、缺乏高频细节等问题, 视觉感知质量较差<sup>[9-10]</sup>。感知质量即图像的视觉效果。相比之下, 感知驱动网络一般基于生成对抗网络 GAN, GAN 在设计中额外引入感知损失, 可以生成边缘清晰、细节丰富的图像<sup>[11]</sup>。然而, GAN 结构中的判别器会引入不稳定因素, 破坏图像结构一致性, 使生成的图像产生不真实的伪影与几何形变<sup>[12]</sup>。此外, 基于 GAN 的网络通常存在训练不稳定、计算量大等问题。通过结合两种网络, 设计一个同时考虑重建精度和感知质量的超分辨率网络, 中和两者的优点, 势必会需要大量的参数和高昂的计算成本, 这对于在临床上使用是不现实的。

知识蒸馏作为一种压缩技术, 可以利用从教师网络中获得的知识来指导学生网络的训练, 将知识从庞

大且计算成本高昂的教师网络转移到一个小型且高效的学生网络<sup>[13]</sup>。近年来, 研究者们提出了一些基于知识蒸馏的自然图像超分辨率方法, 也有个别对脑部 CT、肺部 CT 等医学图像进行了研究<sup>[14]</sup>。这为重建精度和感知质量兼得的 OCT 超分辨率网络的存在提供了可能。

因此, 针对 OCT 图像超分辨率重建网络在重建精度和感知质量上的不平衡问题, 同时为了降低网络的模型复杂度, 更加契合临床使用, 本文提出了一种面向 OCT 图像超分辨率的多教师知识蒸馏网络 MK-OCT。通过对在重建精度和感知质量两个方面性能优秀的教师网络进行学习, 学生网络能够结合教师网络不同的能力, 在平衡优点的同时保持模型的轻量级和高效率。同时, 提出了一种高效通道蒸馏方法 (efficient channel distillation, ECD), 使得学生网络能够从教师网络的中间层中提取丰富的通道注意力信息, 并以损失函数的形式将其传递到学生网络的中间层, 在不增加学生网络参数和计算量的情况下提高模型性能。此外, 对比学习的加入也为超分辨率结果带来了一定提升。实验结果证实了本章提出方法的可行性, 有助于后续 OCT 图像的自动分割、分类等下游任务。

## 2 模型整体框架

知识蒸馏方法基于师生学习范式, 其中通常取较大的教师模型, 并将教师模型的输出作为损失函数中的软目标集成, 以训练通常较小的学生模型。其中, 多教师知识蒸馏采用多个教师网络、单个学生网络的架构, 可以给学生网络传递多种知识。以往大多数使用多教师的知识蒸馏方法在整个提炼过程中简单地对所有教师模型分配了相同的权重。各教师权重的均匀分配使得多教师框架的协调、管理和实施变得简单, 但是, 均匀分配忽视了各种教师模式的优缺点, 为更好地进行知识提炼留下了巨大的未触及空间。理想情况下, 研究者希望根据不同的教师模型在个别情况下的表现, 为不同的训练实例分配不同的权重, 但在实际操作中是较难实现的。同时, 针对 OCT 视网膜图像而言, 纹理层次特征信息往往容易被教师网络的中间层提取, 浅层的学生网络难以通过自身模块或者学习教师网络的输出提取到丰富的细节信息。

为了解决以上问题, 本文提出了一种基于多教师知识蒸馏的超分辨率重建网络 MK-OCT, 模型结构如图 1 所示。MK-OCT 由重建精度教师网络、感知

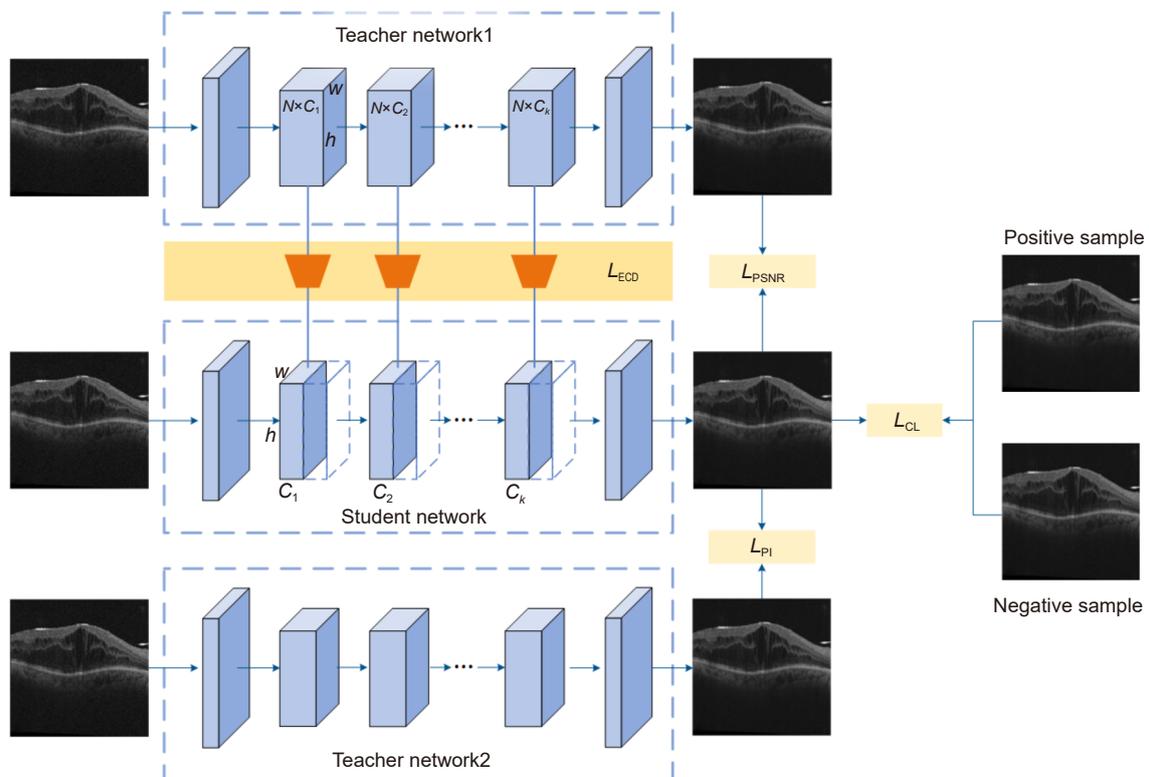


图 1 MK-OCT 整体框架

Fig. 1 Overall framework of MK-OCT

质量教师网络和轻量学生网络三个部分组成, 训练时两个教师网络同时学习从退化的视网膜图像到真实图像的转换, 通过知识的传递, 引导学生网络实现 OCT 视网膜图像超分辨率重建。同时, 为了确保学生网络在简单且浅层的前提下也能保留视网膜图像中关键的纹理和边缘细节, MK-OCT 在学习教师网络的结果之前, 对教师网络的中间层特征信息进行了提取。此外, 通过添加对比学习模块, 限制或者减少了学生网络因为教师网络的不良结果导致性能降低的可能。在训练阶段, 模型所示过程均同时执行; 在测试阶段, 模型只执行学生网络部分。

## 2.1 教师和学生网络设计

在重建精度教师网络和学生网络的设计上, 采用了 MobileNet 架构, 通过将教师网络的通道压缩 4 倍, 并将其卷积滤波器的数量减少到 16 个得到学生网络。在感知质量教师网络的设计中, 考虑到生成对抗网络在生成精细且丰富的细节方面表现优异, 因此大多数感知驱动网络都是基于 GAN 的模型。然而, 基于 GAN 的模型生成的超分辨率图像产生伪影的风险很高, 这可能会对学生网络的训练产生负面影响。同时, 在训练 MK-OCT 时, 每个批次有三个 LR 到 SR 的图像映射, 因此对于教师网络来说, 大多数 GAN 网络计算量过大。

为了保留 GAN 网络在图像感知度方面的良好性能, 并且减少伪影, 同时针对 OCT 图像因散斑噪声和结构相似而无法获得足够精确的细节信息和边缘信息的问题, 本文设计了一个感知教师网络 PASRN 作为  $T_{PI}$ , 其模型结构如图 2 所示。

PASRN 主要是在 ESRGAN<sup>[11]</sup> 的网络结构上做了修改。将 ESRGAN 生成器中的 RRDB 块数量减少到 10 块, 删去鉴别器, 加入 PANet 模块, 同时加入一种边缘损失函数进行训练。其中, PANet 利用特征金字塔和自适应特性池提高图像的细节处理能力, 可以

有效抑制 OCT 图像的噪声, 其结构如图 3 所示。

在 MK-OCT 的训练过程中, 将重建精度教师网络记为  $T_{PSNR}$ , 感知质量教师网络记为  $T_{PI}$ , 学生网络记为  $S_{MK}$ 。输入为低分辨率图像  $I_{LR}$ , 从而得到三个 SR 图像, 分别记为  $I_S$ 、 $I_{SR\_PSNR}$  和  $I_{SR\_PI}$ 。这个过程可以表示为

$$I_S = S_{MK}(I_{LR}), \quad (1)$$

$$I_{SR\_PSNR} = T_{PSNR}(I_{LR}), \quad (2)$$

$$I_{SR\_PI} = T_{PI}(I_{LR}), \quad (3)$$

其中:  $S_{MK}(\cdot)$ 、 $T_{PSNR}(\cdot)$  和  $T_{PI}(\cdot)$  分别表示相应网络的推导过程。

## 2.2 高效通道蒸馏方法 ECD

为了使得学生网络能够更全面地学习到教师网络的有效知识信息, 同时增强 OCT 视网膜图像的纹理细节, 在 ECA-Net<sup>[15]</sup> 的启发下, 本节提出了一种新的高效通道蒸馏方法 ECD, 其处理过程如图 4 所示。学生网络和教师网络分别从各自的特征图中计算出各通道的注意力信息, 随后教师网络通过损失函数来监督学生网络学习各通道的注意力信息, 并将注意力信息传递给学生网络, 从而提升学生网络的性能。对比以往基于特征的知识蒸馏方法中使用的通道蒸馏<sup>[16]</sup>, 本节提出的 ECD 方法能更好地避免维度缩减, 从教师模型中获取丰富的中间层特征信息, 并有效捕获了跨通道交互, 在只涉及少数参数的情况下达到了较好的效果, 降低了模型的复杂度。

具体处理过程: 首先, 使用  $1 \times 1$  卷积对  $S_{MK}$  的特征映射进行维度升级, 使其与  $T_{PSNR}$  保持一致。随后, 利用全局平均池化 (GAP) 将特征图压缩成维度为的特征向量, 通过一维卷积层来完成跨通道间的信息交互。卷积核的大小通过一个函数来自适应变化, 使得通道数较大的层可以更多地进行跨通道交互。自适应函数

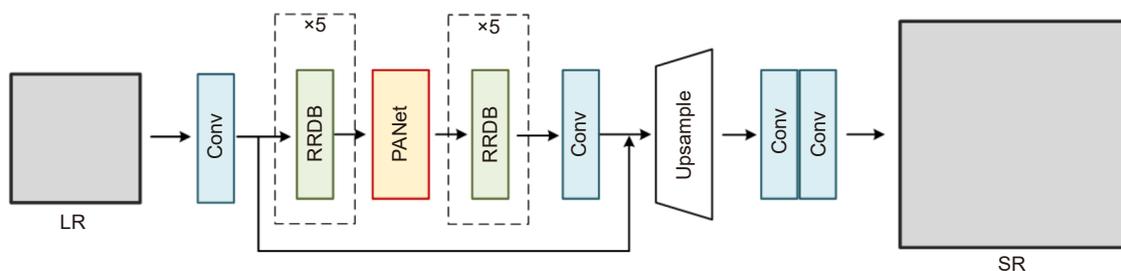


图 2 PASRN 结构  
Fig. 2 Structure of PASRN

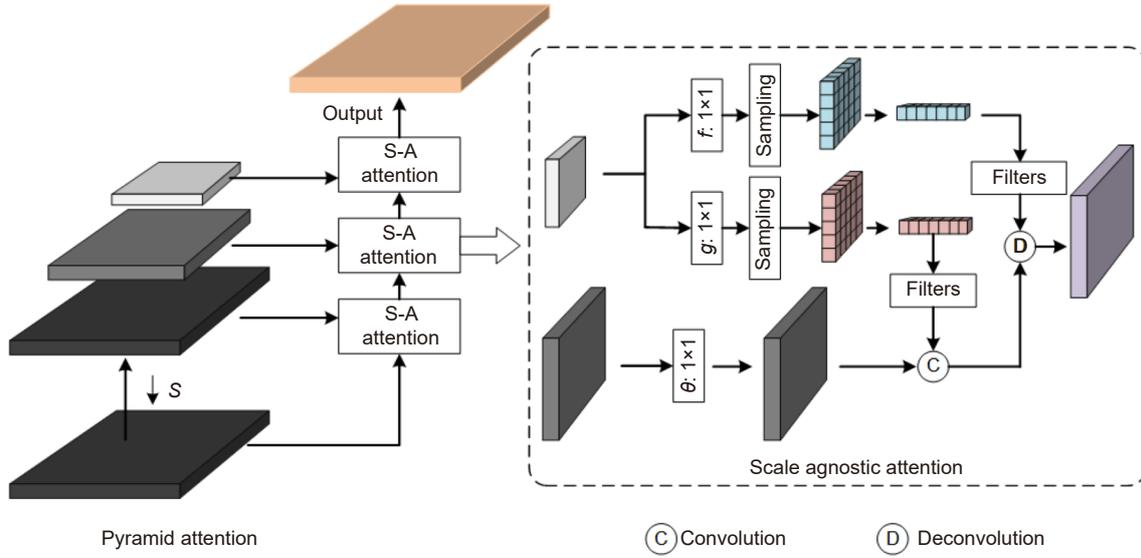


图 3 PANet 结构  
Fig. 3 Structure of PANet

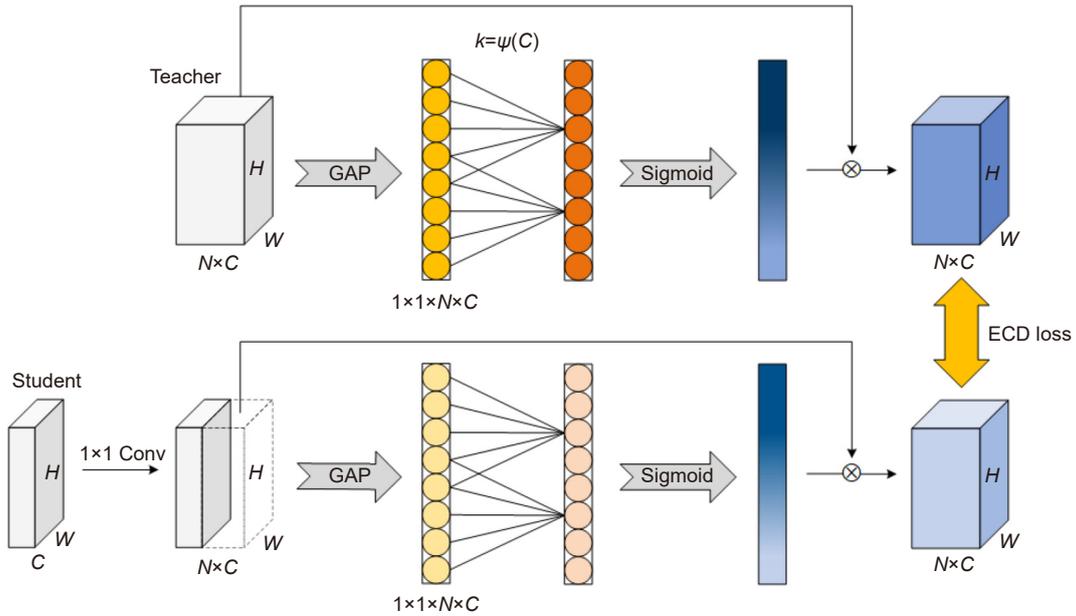


图 4 ECD 模块  
Fig. 4 ECD module

计算表达式如下:

$$k = \Psi(C) = \left\lfloor \frac{\log_2 C}{\gamma} + \frac{b}{\gamma_{\text{odd}}} \right\rfloor, \quad (4)$$

式中:  $\lfloor n \rfloor_{\text{odd}}$  表示与  $n$  最接近的奇数, 同时设置  $\gamma = 2, b = 1$ 。

之后, 通过 Sigmoid 函数得到对于特征图的每个通道的归一化权重, 再将原特征图与之逐通道相乘, 生成加权后的特征图。最后, 计算通道蒸馏损失, 使得  $S_{\text{MK}}$  更有效地获取  $T_{\text{PSNR}}$  各通道的特征信息, ECD

损失函数计算表达式如下:

$$L_{\text{ECD}}(T_{\text{PSNR}}, S_{\text{MK}}) = \frac{\sum_{i=1}^t \sum_{j=1}^c (w_{\text{T}}^{ij} - w_{\text{S}}^{ij})^2}{t \times c}, \quad (5)$$

式中:  $w^{ij}$  表示模型第  $i$  个特征映射的第  $j$  个通道的注意力权重,  $t$  表示特征映射的个数,  $c$  表示通道的个数。

### 2.3 对比学习

对比学习 (Contrastive learning, CL) 是一种自监督学习方法, 其核心是通过计算样本特征间的距离, 将

锚点拉近到接近正样本的位置, 同时将其推离负样本, 其在本模型中的应用如图 5 所示。相比于像素细节, 对比学习更多地关注抽象的语义信息, 并且对比学习的另一个优势在于更加容易优化。在对比学习的设计中, 有两个需要考虑的关键: 一是设计构建合适的正负样本, 二是寻找潜在特征空间来进行样本的比较。在本文中, 对比学习可以提供有上限和下限的外部知识, 为 OCT 图像超分辨率任务减少优化空间, 从而进一步提高学生网络的性能。此外, 在知识蒸馏中, 当教师网络的结果有一定偏差时, 如果仅仅依靠于教师网络的结果, 学生网络的性能也会受到影响, 因此对比学习的添加可以用于限制或者减少教师网络的不良结果带来的问题。

本文设计选取真实 HR 图像为正样本, 记为  $I_P$ ; 低分辨率图像双三次上采样后的图像为负样本, 记为  $I_N$ 。选择相同的固定预训练模型 VGG19 中的公共中间层特征为潜在特征空间, 计算对比损失, 表达式如下:

$$L_{CL} = \sum_{i=1}^N \frac{d(\varphi_i(I_S), \varphi_i(I_P))}{d(\varphi_i(I_S), \varphi_i(I_N))}, \quad (6)$$

其中:  $d(x,y)$ 表示  $x$  和  $y$  之间的距离,  $\varphi_i$ 表示从固定预训练模型中提取的第  $i$  个隐藏特征。

## 2.4 损失函数

MK-OCT 通过四个损失函数对生成的 SR 图像做损失计算。首先, 为了学习  $T_{PSNR}$  在图像重建精度上的能力, 计算  $I_S$  和  $I_{SR\_PSNR}$  之间的平均绝对误差 MAE, 即  $L_1$  损失, 表达式如下:

$$L_1(I_{SR}, I_{SR\_PSNR}) = \|I_{SR} - I_{SR\_PSNR}\|_1. \quad (7)$$

同时, 为了学习  $T_{PQ}$  在感知质量方面的能力, 防止重建图像过于平滑, 本文使用预训练的 19 层 VGG 网络计算  $I_S$  和  $I_{SR\_PI}$  之间的感知损失, 表达式如下:

$$L_P(I_{SR}, I_{SR\_PI}) = \|VGG(I_{SR}) - VGG(I_{SR\_PI})\|_1, \quad (8)$$

因此, MK-OCT 的总损失函数可以表示为

$$I_{MK-OCT} = \lambda_1 L_1 + \lambda_2 L_P + \lambda_3 L_{CL} + L_{ECD}, \quad (9)$$

其中:  $\lambda_1$ 、 $\lambda_2$  和  $\lambda_3$  分别表示平衡逐像素重建误差、感知损失和对比损失的超参数。

## 3 实验

### 3.1 实验设置

本文所使用两个真实数据集, 分别由绵阳市第三人民医院和福建医科大学附属协和医院提供, 包含成对的含噪和干净的 OCT 图像, 均为由临床使用的德

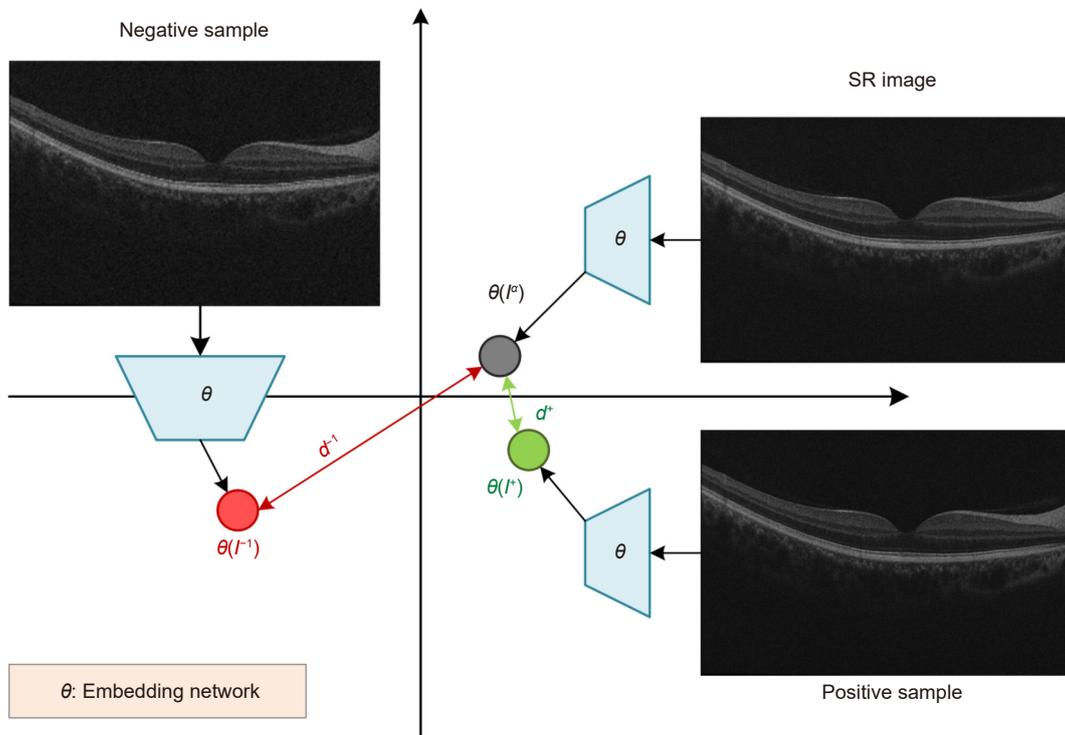


图 5 对比学习示意图  
Fig. 5 Contrastive learning

国 ZEISS 视网膜光学相干断层扫描仪获取的真实图像, 图像尺寸为 1055×703。为了满足实验需求, 同时提高模型的训练质量, 实验中裁去原始图像的白边和外围部分的模糊结构, 裁剪后的尺寸为 960×640, 作为 HR 图像。对原始含噪图像进行 4 倍下采样来模拟临床实践中的低采样率, 生成×4 的含噪 LR 图像。其中, 两个数据集均选取 1000 对图像作为训练集, 80 对图像作为验证集, 120 对图像作为测试集。训练集、验证集和测试集的图像均没有交集。

为增加网络的训练数据, 提高模型泛化能力, 实验中采取两个措施增强数据集: 1) 将图像旋转 90°、水平翻转; 2) 采用了 Yoo 等人提出的适用于真实图像超分辨率的混合数据增强方法 MoA<sup>[17]</sup>。实验中采用 Adam (adaptive moment estimation, 自适应矩估计) 作为优化器, 其具体参数设置为  $\beta_1=0.9$ 、 $\beta_2=0.999$ 、 $\epsilon=10^{-8}$ 。训练过程中, 图像批处理大小设置为 8, 初始学习率为  $1 \times 10^{-3}$ , 每 80 个周期学习率衰减一半。

### 3.2 评价指标

为客观地与其他经典图像超分辨率算法进行重建精度和感知质量的比较, 本文采用峰值信噪比 (peak signal-to-noise ratio, PSNR)、结构相似性 (structural similarity, SSIM)、学习感知图像块相似度 (learned perceptual image patch similarity, LPIPS) 和图像感知质量指标 (perceptual index, PI) 来客观描述超分辨率图像的重建质量。其中, PSNR、SSIM、LPIPS 为全参考指标, PI 为非参考指标。PSNR 和 SSIM 用于衡量重建保真度, LPIPS 和 PI 用于评估感知质量。

PSNR 用于衡量图像的重建质量, 该值越大, 意味着重建效果越好。计算公式如下:

$$R_{\text{PSN}} = 10 \times \log_{10} \left( \frac{1}{H \times W} \|I_s - I_H\|_2^2 \right)^{-1}, \quad (10)$$

其中:  $I_s$  为重建获得的 SR 图像,  $I_H$  为原始 HR 图像。

SSIM 是一种衡量两幅图像相似度的指标, 在评价图像质量上更能符合人类的视觉特性。该值越大, 意味着重建图像与原始图像误差越小。计算公式如下:

$$SSIM(x, y) = [l(x, y)]^\alpha \cdot [c(x, y)]^\beta \cdot [s(x, y)]^\gamma, \quad (11)$$

其中:  $l(x, y)$  为亮度比较,  $c(x, y)$  为对比度比较,  $s(x, y)$  为结构比较。实际应用中  $\alpha = \beta = \gamma = 1$ 。

LPIPS 是一种基于学习的感知图像补丁相似性指标, 用于评估图像的感知质量, 通过学习一个神经网络模型来近似人类感知的视觉相似性。LPIPS 的得分范围通常是 0 到 1 之间, 数值越小表明图像的感知质量越高。其公式如下:

$$d(x, x_0) = \sum_l \frac{1}{H_l W_l} \sum_{h, w} \left\| w_l \odot (\hat{y}_{hw}^l - \hat{y}_{0hw}^l) \right\|_2^2, \quad (12)$$

其中:  $x$  为 SR 图像,  $x_0$  为 HR 图像,  $d$  为两者之间的距离。将 SR 图像和 HR 图像送入神经网络中进行特征提取, 然后进行激活后归一化处理, 记为  $\vec{y}^l$ ,  $\vec{y}_0^l \in R^{H_l \times W_l \times C_l}$ 。本文选择 VGG 网络进行特征提取。

PI 值代表的是一幅图像的主观感知质量, 往往 PI 值越低, 图像的主观视觉效果越佳, 则图像的感知质量越好。其公式如下:

$$PI = \frac{1}{2} ((10 - Ma) + NIQE), \quad (13)$$

其中:  $Ma$  代表马氏评分,  $NIQE$  为图像评价质量。

## 4 数据与结果分析

### 4.1 定量分析

为验证本文提出的模型在 OCT 图像超分辨率重建上的优越性, 将 MK-OCT 与×4 尺度下的几种轻量级 SR 网络进行了比较, 包括 SRCNN<sup>[18]</sup>、IMDN<sup>[19]</sup>、RFDN<sup>[20]</sup> 和 CSD<sup>[21]</sup>。SRCNN 是提出时间较早的经典超分辨率重建网络, CSD 是使用知识蒸馏的超分辨率重建网络, IMDN 和 RFDN 是近年来提出的轻量级超分辨率重建网络。其中, CSD 采用自蒸馏模型, 其余采用 LR-HR 配对图像进行训练。

表 1 显示了在两个数据集上, 四个指标的比较结果。在重建精度方面, 即 PSNR 和 SSIM, 除了双三次 (Bicubic) 插值和 SRCNN, 其他模型的性能均比较优秀, MK-OCT 略优于其他模型, 表明其重建精度更好。这是因为 MK-OCT 中使用的学生网络 BSRN 为卷积神经网络, 专注于对局部特征的提取, 而作为重建精度教师的 ELAN 使用自注意力机制, 能够建立全局的依赖关系, 获取更多上下文信息, 使得 BSRN 能够学习到 ELAN 的优点, 弥补自身网络的不足, 进一步提高重建精度。

在感知质量方面, 即 LPIPS 和 PI, 可以明显看出 MK-OCT 在这几个轻量级网络中得到了最好的效果, 低至 0.137, 更符合人眼的视觉感知, 这是因为其学习了面向感知质量的教师网络所传递的知识信息。

表 1 各种超分辨率模型在 4 倍重建后的平均指标

Table 1 Average performance of various super-resolution models after x4 reconstruction

Method	Size /MB	FLOPs /G	Dataset 1				Dataset 2			
			PSNR	SSIM	LPIPS	PI	PSNR	SSIM	LPIPS	PI
Bicubic	-	-	28.12	0.7811	0.412	6.795	28.43	0.7730	0.422	6.579
SRCNN	0.2	0.23	28.59	0.8003	0.404	6.355	28.79	0.7986	0.398	6.297
CSD	12.16	122.1	30.95	0.8142	0.310	5.677	30.90	0.8119	0.327	5.802
IMDN	2.65	41.9	31.24	0.8217	0.226	5.553	31.21	0.8220	0.230	5.608
RFDN	1.59	32.0	31.67	0.8262	0.220	5.217	31.78	0.8217	0.217	5.139
MK-OCT (Ours)	1.41	29.8	<b>32.93</b>	<b>0.8460</b>	<b>0.149</b>	<b>4.521</b>	<b>32.90</b>	<b>0.8443</b>	<b>0.143</b>	<b>4.443</b>

同时, 两个数据集相比较可以看出, MK-OCT 在数据集 2 上测试所生成的图像感知质量更好, 推测是因为数据集 2 中的图像多为病变视网膜 OCT 图像, 纹理细节较多, 因此 MK-OCT 在感知质量上的进步更明显。此外, 除了 SRCNN, MK-OCT 模型最小且浮点计算次数最少, 相比其他模型更适合在临床终端上部署使用。

#### 4.2 定性分析

本节实验将对重建后的图像进行定性分析, 分别选取了数据集 1 和数据集 2 测试集中的一张 OCT 图像为例, 并突出放大图像的重点部分, 将本文提出的模型 MK-OCT 与 SRCNN、CSD、IMDN 和 RFDN 进行视觉效果上的直观对比, 结果如图 6 所示。除了 HR 图像和经由 MK-OCT 处理后的 SR 图像, 为了更加直观地看出各个模型重建后的效果, 图 6 提供了有关眼部疾病分析的局部重建图像。

从图像清晰度和对细节的还原度来看, 与原始的 HR 图像相比, SRCNN 虽然模型计算量小, 但重建结果最为模糊, 整体清晰度低; 经 CSD 重建后的图像在整体清晰度上有所提高, 但是边缘依旧缺失明显; IMDN 和 RFDN 在空间域中展示了较为清晰的纹理, 但是存在伪影, 同时可以明显看出 IMDN 重建后的图像有过度平滑的缺点, 这导致视网膜结构呈现出粘连, 精细部分细节缺失, 极易误导医生判断。相比之下, MK-OCT 学习了两个教师网络传递的信息, 重建图像不仅整体清晰度高, 同时具有更好的感知质量, 纹理更细, 边缘更锐利, 同时也没有明显的噪声和伪影问题, 在最小的计算量下更有效地还原出了图像的高频信息。

从视网膜的生理结构分层上来看, MK-OCT 对于外界膜的重建最为清晰连贯, 并且对于内层视网膜

的结构分层较为明确, 例如神经纤维层、神经节细胞层、内丛状层、内颗粒层等, 都能一目了然地分辨。这能够帮助医生更加清楚地观察视网膜的病理特点, 为后续病理分析提供更好的选择。

#### 4.3 消融实验

为了充分证明 MK-OCT 模型的有效性, 本文进行了消融实验。在四种条件下对  $S_{MK}$  进行 4 倍超分辨率重建训练, 分别是直接使用 LR-HR 图像对训练的 HR- $S_{MK}$ 、重建精度单教师知识蒸馏的  $T_{PSNR}$ 、感知质量单教师知识蒸馏的  $T_{PI}$  和不使用对比学习的 None-CL。表 2 为四种实验设计的消融实验结果, 其中每个类别的最佳结果以粗体突出显示。

从 PSNR 及 SSIM 的指标上看, 表现最好的是重建精度单教师知识蒸馏的模型; 从 LPIPS 和 PI 来看, 性能最佳的是感知质量单教师知识蒸馏的模型。两个单教师知识蒸馏下的模型分别在其教师网络擅长的领域表现优异, 在其他方面表现一般, 这证明了知识蒸馏的有效性, 同时也进一步解释了多教师知识蒸馏性能较好的原因。同时, 从表 1 和表 2 的对比可以看出, 相比于直接从 HR 图像中学习, 从单教师网络中学习的学生网络性能更加优秀, 这是因为教师网络中传递的知识可学习性更强。另外, 与进行了对比学习的模型 MK-OCT 相比, 未进行对比学习的模型性能略有下降, 因此也可证明对比学习的加入为 MK-OCT 的性能带来了一定的提高。

#### 4.4 泛化性能

模型的泛化能力是指模型在未曾见过的数据上的表现能力, 也就是模型对于新的数据的适应能力。眼科临床中, 不同医院所使用的 OCT 设备不同, 可能导致产生的 OCT 图像有所区别, 因此, 一个好的模型应该具有良好的泛化能力, 以适应不同的 OCT 设备。

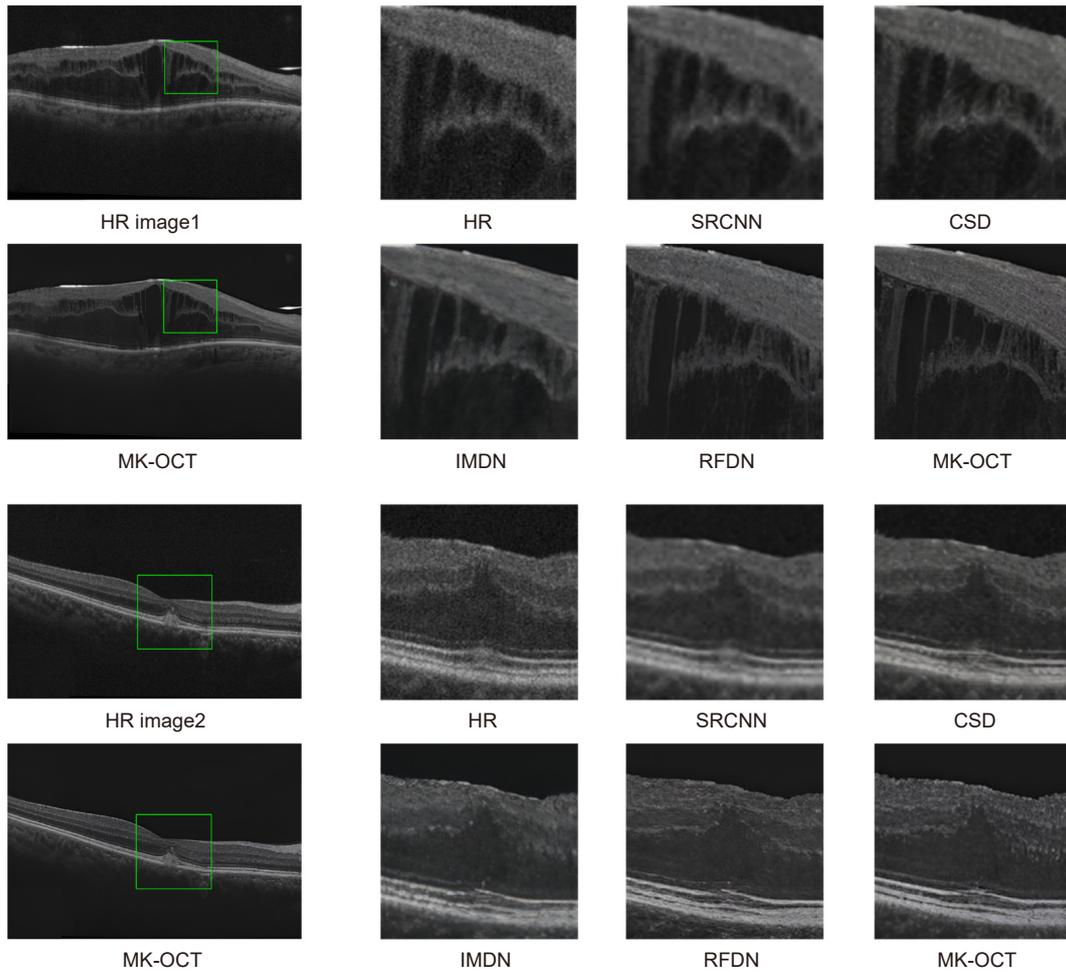


图 6 超分辨率重建结果

Fig. 6 Results of super-resolution reconstruction

表 2 不同条件的学生网络在 4 倍重建后的定量评估

Table 2 Quantitative evaluation of student networks under different conditions after x4 reconstruction

Dataset	Metric	HR-S <sub>MK</sub>	Single-teacher		None-CL
			$T_{PSNR}$	$T_{PI}$	
Dataset 1	PSNR	31.27	<b>32.88</b>	32.77	<b>32.88</b>
	SSIM	0.8238	<b>0.8459</b>	0.8396	0.8457
	LPIPS	0.230	0.217	<b>0.142</b>	0.150
	PI	5.593	5.440	<b>4.457</b>	4.608
Dataset 2	PSNR	31.33	<b>32.87</b>	32.81	32.86
	SSIM	0.8178	<b>0.8424</b>	0.8411	0.8420
	LPIPS	0.214	0.209	<b>0.140</b>	0.148
	PI	5.146	5.129	<b>4.561</b>	4.670

为验证 MK-OCT 的泛化能力, 本小节从第四届 MICCAI 眼科医学图像分析研讨会<sup>[22]</sup>的 OCT 视网膜图像中, 随机抽取了 100 对图像作为测试集, 利用这些算法对测试集进行了 2 倍和 4 倍的超分辨率重建,

同时采用 PSNR 和 PI 两个指标来量化比较, 并将结果列于表 3 中。相较于其他算法, MK-OCT 在 PSNR、PI 上均有优势, 充分证明了 MK-OCT 模型具有良好的泛化能力。

表 3 新数据集上各种超分辨率模型的平均 PSNR 和 PI 值

Table 3 Average PSNR and PI values of various super-resolution models after reconstruction

Method	PSNR		PI	
	×2	×4	×2	×4
SRCNN	33.67	28.79	4.667	6.033
CSD	34.22	29.98	4.109	5.820
IMDN	35.90	31.06	4.233	5.709
RFDN	35.89	31.77	4.059	5.455
MK-OCT (Ours)	<b>36.20</b>	<b>32.58</b>	<b>3.979</b>	<b>5.103</b>

## 5 结论

为了解决深度学习方法因其计算复杂度难以在设备较落后的社区医院、卫生院中应用的问题, 同时平衡超分辨率网络的重建精度和感知质量, 本文提出了针对 OCT 图像超分辨率的多教师知识蒸馏网络 MK-OCT, 采用在重建精度和感知质量两方面分别擅长的教师网络来训练一个平衡、轻量级和高效的学生网络。其中, 提出了一种新的高效通道蒸馏方法 ECD, 使得学生网络能从教师网络的中间层获取丰富的特征信息。同时, 考虑到模型整体计算量, 提出了由 ESRGAN 改进而来的 PASRN 作为感知驱动的教师网络, 规避 GAN 网络固有风险的同时, 保证了其在感知质量方面的优点。此外, MK-OCT 将知识蒸馏和对比学习相结合, 减少了优化空间, 进一步提升重建质量。实验证明, 本文所提模型可以有效地提高 OCT 视网膜图像的分辨率, 与 SRCNN、CSD、IMDN 和 RFDN 等轻量级超分辨率模型相比, MK-OCT 在复杂度更低的前提下, 重建精度和感知质量上都有一定程度的提高, 尤其是在感知质量上的提高比较明显, 并且在主观视觉上也有明显改善。最后, 经过细微调整, 任何其他现有的训练模型都可以作为教师或学生网络放到 MK-OCT 框架中进行训练使用。MK-OCT 也可以拓展到其他相关任务, 如 OCT 图像阴影修复以及其他医学图像任务。

**利益冲突:** 所有作者声明无利益冲突

## 参考文献

- [1] Lu D X, Fang W H, Li Y Y, et al. Optical coherence tomography: principles and recent developments[J]. *Chin Opt*, 2020, 13(5): 919–935.  
陆冬筱, 房文汇, 李玉瑶, 等. 光学相干层析成像技术原理及研究进展[J]. *中国光学*, 2020, 13(5): 919–935.
- [2] Huang Y Q, Lu Z X, Shao Z M, et al. Simultaneous denoising and super-resolution of optical coherence tomography images based on generative adversarial network[J]. *Opt Express*, 2019, 27(9): 12289–12307.
- [3] Das V, Dandapat S, Bora P K. Unsupervised super-resolution of OCT images using generative adversarial network for improved age-related macular degeneration diagnosis[J]. *IEEE Sensors J*, 2020, 20(15): 8746–8756.
- [4] Qiu B, You Y F, Huang Z Y, et al. N2NSR - OCT: simultaneous denoising and super-resolution in optical coherence tomography images using semisupervised deep learning[J]. *J Biophotonics*, 2021, 14(1): e202000282.
- [5] Lu Y Q, Chen M H, Qin K B, et al. Super-resolution reconstruction of OCT image based on pyramid long-range transformer[J]. *Chin J Lasers*, 2023, 50(15): 1507107.  
芦焱琦, 陈明惠, 秦楷博, 等. 基于金字塔长程 Transformer 的 OCT 图像超分辨率重建[J]. *中国激光*, 2023, 50(15): 1507107.
- [6] Ke S T, Chen M H, Zheng Z X, et al. Super-resolution reconstruction of optical coherence tomography retinal images by generating adversarial network[J]. *Chin J Lasers*, 2022, 49(15): 1507203.  
柯舒婷, 陈明惠, 郑泽希, 等. 生成对抗网络对 OCT 视网膜图像的超分辨率重建[J]. *中国激光*, 2022, 49(15): 1507203.
- [7] Ma Y H, Chen X J, Zhu W F, et al. Speckle noise reduction in optical coherence tomography images based on edge-sensitive cGAN[J]. *Biomed Opt Express*, 2018, 9(11): 5129–5146.
- [8] Wang R G, Lei H, Yang J. Self-similarity enhancement network for image super-resolution[J]. *Opto-Electron Eng*, 2022, 49(5): 210382.  
汪荣贵, 雷辉, 杨娟, 等. 基于自相似特征增强网络结构的图像超分辨率重建[J]. *光电工程*, 2022, 49(5): 210382.
- [9] Ma C, Rao Y M, Cheng Y, et al. Structure-preserving super resolution with gradient guidance[C]//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 7766–7775.  
<https://doi.org/10.1109/CVPR42600.2020.00779>.
- [10] Park S J, Son H, Cho S, et al. SRFeat: single image super-resolution with feature discrimination[C]//*Proceedings of the 15th European Conference on Computer Vision*, 2018: 455–471. [https://doi.org/10.1007/978-3-030-01270-0\\_27](https://doi.org/10.1007/978-3-030-01270-0_27).
- [11] Wang X T, Yu K, Wu S X, et al. ESRGAN: enhanced super-resolution generative adversarial networks[C]//*Proceedings of the 15th European Conference on Computer Vision*, 2018: 63–79. [https://doi.org/10.1007/978-3-030-11021-5\\_5](https://doi.org/10.1007/978-3-030-11021-5_5).
- [12] Yao G Q, Li Z, Bhanu B, et al. MTKDSR: multi-teacher knowledge distillation for super resolution image reconstruction[C]//*Proceedings of the 2022 26th International Conference on Pattern Recognition (ICPR)*, 2022: 352–358. <https://doi.org/10.1109/ICPR56361.2022.9956250>.

- [13] Shu C Y, Liu Y F, Gao J F, et al. Channel-wise knowledge distillation for dense prediction[C]//*Proceedings of 2021 IEEE/CVF International Conference on Computer Vision*, 2021: 5291–5300. <https://doi.org/10.1109/ICCV48922.2021.00526>.
- [14] Zhao T L, Hu L, Zhang Y M, et al. Super-resolution network with information distillation and multi-scale attention for medical CT image[J]. *Sensors*, 2021, 21(20): 6870.
- [15] Wang Q L, Wu B G, Zhu P F, et al. ECA-Net: efficient channel attention for deep convolutional neural networks[C]//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 11531–11539. <https://doi.org/10.1109/CVPR42600.2020.01155>.
- [16] Zhou Z D, Zhuge C R, Guan X W, et al. Channel distillation: channel-wise attention for knowledge distillation[Z]. arXiv: 2006.01683, 2020. <https://doi.org/10.48550/arXiv.2006.01683>.
- [17] Yoo J, Ahn N, Sohn K A. Rethinking data augmentation for image super-resolution: a comprehensive analysis and a new strategy[C]//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 8372–8381. <https://doi.org/10.1109/CVPR42600.2020.00840>.
- [18] Dong C, Loy C C, He K M, et al. Learning a deep convolutional network for image super-resolution[C]//*Proceedings of the 13th European Conference on Computer Vision*, 2014: 184–199. [https://doi.org/10.1007/978-3-319-10593-2\\_13](https://doi.org/10.1007/978-3-319-10593-2_13).
- [19] Hui Z, Gao X B, Yang Y C, et al. Lightweight image super-resolution with information multi-distillation network[C]//*Proceedings of the 27th ACM International Conference on Multimedia*, 2019: 2024–2032. <https://doi.org/10.1145/3343031.3351084>.
- [20] Liu J, Tang J, Wu G S. Residual feature distillation network for lightweight image super-resolution[C]//*Proceedings of the 16th European Conference on Computer Vision*, 2020: 41–55. [https://doi.org/10.1007/978-3-030-67070-2\\_2](https://doi.org/10.1007/978-3-030-67070-2_2).
- [21] Wang Y B, Lin S H, Qu Y Y, et al. Towards compact single image super-resolution via contrastive self-distillation[C]//*Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, 2021: 1122–1128. <https://doi.org/10.24963/ijcai.2021/155>.
- [22] Bogunović H, Venhuizen F, Klimscha S, et al. RETOUCH: the retinal OCT fluid detection and segmentation benchmark and challenge[J]. *IEEE Trans Med Imaging*, 2019, 38(8): 1858–1874.

## 作者简介



【通信作者】陈明惠(1981-), 女, 博士, 副教授, 从博士阶段开始近十年一直投入在医用光学相干层析(OCT)实时成像技术的研究, 研制扫频激光等装置和器械, 应用在光学相干层析成像系统。

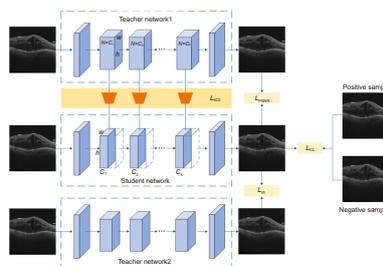
E-mail: [cmhui.43@163.com](mailto:cmhui.43@163.com)



扫描二维码, 获取PDF全文

# Super-resolution reconstruction of retinal OCT image using multi-teacher knowledge distillation network

Chen Minghui<sup>1\*</sup>, Lu Yanqi<sup>1</sup>, Yang Wenyi<sup>1</sup>, Wang Yuanzhu<sup>2</sup>, Shao Yi<sup>3</sup>



Overall framework of MK-OCT

**Overview:** Optical coherence technology (OCT), which is widely used in the diagnosis of ophthalmic diseases, can reconstruct three-dimensional cross-sectional images inside biological tissues through the mutual interference of weakly coherent light. However, due to the inevitable scattering of weakly coherent light when it enters the tissue, there is speckle noise in the OCT retinal image, which covers up the subtle and very important details in the image. Secondly, unconscious movements such as eye movements (drift, tremors, and micro jumps), head movements, and cardiopulmonary system during the image acquisition process can lead to artifacts in OCT images, affecting clinical diagnosis and interfering with subsequent automated analysis of images. To solve the problem of existing OCT super-resolution networks being solely focused on reconstruction accuracy and perceptual quality, reduce the model complexity of the network, and be more suitable for clinical applications, this paper proposes a multi teacher knowledge distillation network MK-OCT for OCT image super-resolution. Through knowledge distillation, the student network can combine the different abilities of the teacher network to achieve balance, lightweight, and efficiency. At the same time, an efficient channel distillation method ECD was proposed, which enables the student network to extract rich channel attention information from the middle layer of the teacher network and transmit it to the middle layer of the student network in the form of a loss function, improving model performance without increasing the parameters and computational complexity of the student network. During the training process, both the student network and the teacher network use low-resolution images as input, and after the three networks respectively obtain reconstructed images, different loss functions are used to calculate the loss between the output images of each network. This allows the student network to simultaneously learn both reconstruction accuracy and perceptual quality from the two teacher networks. In addition, the student network additionally uses contrastive learning, which can provide external knowledge with upper and lower bounds, reducing the optimization space for the OCT image super-resolution task, thereby further improving the performance of the student network. We compared our model to five classic lightweight super-resolution reconstruction models, namely SRCNN, CSD, IMDN, and RFDN. Experiments have verified the effectiveness and superiority of MK-OCT in OCT image super-resolution reconstruction. At the same time, our research group also conducted ablation experiments, which further confirmed the effectiveness of multi teacher knowledge distillation. The generalization performance experiment also proves that the MK-OCT model has a good generalization ability.

Chen M H, Lu Y Q, Yang W Y, et al. Super-resolution reconstruction of retinal OCT image using multi-teacher knowledge distillation network[J]. *Opto-Electron Eng*, 2024, 51(7): 240114; DOI: 10.12086/oe.2024.240114

Foundation item: Project supported by Shanghai Science and Technology Commission's Industry University Research Medical Project (15DZ1940400)

<sup>1</sup>Shanghai Engineering Research Center of Interventional Medical, Shanghai Institute for Interventional Medical Devices, School of Health Sciences and Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China; <sup>2</sup>Shanghai Raykeen Laser Technology Co., Ltd., Shanghai 200120, China; <sup>3</sup>Shanghai General Hospital, Shanghai 200080, China

\* E-mail: cmhui.43@163.com